

Metafrontier efficiency analysis with convex and non-convex metatechnologies by stochastic nonparametric envelopment of data

Mohsen Afsharian

*Department of Business Sciences, Technische Universität Braunschweig, Fallersleber-Tor-Wall 23,
38100 Braunschweig, Germany, m.afsharian@tu-bs.de*

Abstract. This paper suggests how stochastic nonparametric envelopment of data (StoNED) can be extended as an estimator in the metafrontier efficiency analysis. Both convex and non-convex metatechnologies are formed and a semi-nonparametric estimation technique for the corresponding metafrontiers is developed. Remaining consistent with the metafrontier theory, the resulting estimated metafrontiers always envelope the estimated group frontiers.

Keywords: Metafrontier analysis; Convex and non-convex metatechnologies; Stochastic nonparametric envelopment of data (StoNED).

JEL-classification: C6, C13, C14

1. Introduction

Introduced by Hayami and Ruttan (1970) and operationalized by Battese et al. (2004) as well as O'Donnell et al. (2008), the metafrontier efficiency analysis has widely been applied in various environments. Within this approach, firms may be classified into different groups, each operating under the same technology. The metatechnology then includes all production possibilities that can in principle be achieved by assuming that the operating environment for firms can be changed.

There are two different ways in which the metatechnology could be defined: *convex and non-convex metatechnologies*. Whereas the former is the convex hull of all group technologies, the latter is formed as a pure union of the group technologies. The analytical framework necessary for the application of these metatechnologies has been given by O'Donnell et al. (2008). An updated overview of the metafrontier methodology can be found in Kerstens et al. (2015).

This paper suggests how stochastic nonparametric envelopment of data (StoNED)¹ can be extended as an estimator in the metafrontier efficiency analysis. Both convex² and non-convex metatechnologies are formed and a semi-nonparametric estimation technique for the corresponding metafrontiers is developed. Remaining consistent with the metafrontier theory, the resulting estimated metafrontiers always envelope the estimated group frontiers.

StoNED-based approaches require generally solving a series of non-linear problems. A further aim of this paper is to formulate equivalent linear and mixed integer linear programs for the computation of the meta-efficiencies within our approach. This significantly simplifies the implementation of the method.

2. Preliminaries

We assume that there exist n firms which are partitioned into $G > 1$ distinct groups. Let each group $g \in \mathbf{G} = \{1, \dots, G\}$ include δ_g observations, which use inputs $\mathbf{x}_j^g = (x_{1j}^g, x_{2j}^g, \dots, x_{mj}^g) \in \mathfrak{R}_+^m$ to produce a

¹ StoNED, proposed first by Kuosmanen (2006), combines the nonparametric, piece-wise linear style of data envelopment analysis (DEA) with the stochastic style of stochastic frontier analysis (SFA) treatment of inefficiency and noise. See also Kuosmanen et al. (2015) for an updated technical overview.

² In the context of measuring the productivity of dairy farms, Sipiläinen et al. (2008) presented a sketch of how StoNED may be used in the analysis where the metatechnology is convex. In the design of our convex approach, we have benefited from their discussions.

single output $y_j^g \in \mathbb{R}_+$. We view each group of firms as operating within a different technology T^g , which can be characterized by the output set $S^g(\mathbf{x}) = \{y \mid (\mathbf{x}, y) \in T^g\}$. We refer to the boundary of these sets as *group frontiers*, which indicate the maximum output that can be produced with the inputs in group $g \in \mathbf{G}$. The within-group efficiencies (indicated in the following by eff_j^g) can be measured against the group frontier $g \in \mathbf{G}$ by the output distance function $\hat{D}^g(\mathbf{x}_j^g, y_j^g) = \min \left\{ \theta > 0 \mid (\mathbf{x}_j^g, y_j^g / \theta) \in T^g \right\}$.

The metatechnology for the group of technologies T^g , $g \in \mathbf{G}$, is defined as

$$T^M = T^1 \cup T^2 \cup \dots \cup T^G, \quad (1)$$

whose boundary is referred to as *metafrontier*. The meta-efficiencies against this metafrontier – represented in the following by eff_j^M – can be measured by the output metadistance function $\hat{D}^M(\mathbf{x}_j^g, y_j^g) = \min \left\{ \theta > 0 \mid (\mathbf{x}_j^g, y_j^g / \theta) \in T^M \right\}$.

According to (1), irrespective of the properties of sets $S^g(\mathbf{x})$, we should have $eff_j^M \leq eff_j^g$ for all $g \in \mathbf{G}$. The gap between the within-group efficiency and the meta-efficiency is also computed by the metatechnology ratio $MTR(\mathbf{x}_j^g, y_j^g) = eff_j^M / eff_j^g$.

3. Estimation of within-group efficiencies

Let each group technology T^g be represented by a group frontier production function $f^g(\mathbf{x})$ such that $T^g = \{(\mathbf{x}, y) \mid y \leq f^g(\mathbf{x})\}$. We assume that these functions $f^g(\mathbf{x})$ are continuous, monotonic increasing, and concave. As highlighted by Kuosmanen et al. (2015), this is equivalent to stating that each T^g satisfies the classic DEA assumptions of free disposability and convexity. In contrast to SFA, no specific functional form for $f^g(\mathbf{x})$ is needed.

In each group $g \in \mathbf{G}$, the observed output $y_j^g \in \mathfrak{R}_+$ of firm j may differ from $f^g(\mathbf{x}_j^g)$ due to inefficiency $u_j^g > 0$ and noise v_j^g , expressed by $y_j^g = f^g(\mathbf{x}_j^g) + \varepsilon_j^g$ where $\varepsilon_j^g = v_j^g - u_j^g$. The inefficiency term $u_j^g \sim N^+(\mu^g, \sigma_{gu}^2)$ and the stochastic noise term $v_j^g \sim N(0, \sigma_{gv}^2)$ are assumed to be statistically independent of each other as well as of inputs \mathbf{x}_j^g . Within the StoNED method, the shape of the functions f^g , $g \in \mathbf{G}$, is estimated by the following quadratic programming (QP) model:

$$\begin{aligned} \min \quad & \sum_{j=1}^{\delta_g} (\varepsilon_j^g)^2 \\ \text{s.t.} \quad & y_j^g = \alpha_j^g + \boldsymbol{\beta}_j^g \mathbf{x}_j^g + \varepsilon_j^g, \quad \forall j \\ & \alpha_j^g + \boldsymbol{\beta}_j^g \mathbf{x}_j^g \leq \alpha_h^g + \boldsymbol{\beta}_h^g \mathbf{x}_j^g, \quad \forall j, \forall h \\ & \boldsymbol{\beta}_j^g \geq 0, \quad \forall j, \end{aligned} \quad (2)$$

where α_j^g and vector $\boldsymbol{\beta}_j^g$ define the intercept and slope parameters of the tangent hyperplanes that characterize the underlying true functions in each group $g \in \mathbf{G}$.

Model (2) provides the composite residuals ε_j^g , which consist of error and inefficiency. Extracting these residuals, one can use, e.g., the method of moments introduced by Aigner et al. (1977) to determine the expected value of ε_j^g which can be used to shift the estimates of the production function upwards in order to define the production frontier for each group. The conditional expected value of efficiency for each firm, represented by $E(u_j^g | \varepsilon_j^g)$, can also be estimated by the method developed by Jondrow et al. (1982). The within-group efficiencies eff_j^g can then be computed as $eff_j^g = 1 - E(u_j^g | \varepsilon_j^g) / f^g(\mathbf{x}_j^g)$. See, e.g., Kuosmanen (2015) for more details.

4. Estimation of meta-efficiencies

One may apply the same method in Section 3 on the data set that includes all firms in all groups to estimate the meta-efficiencies. This naive approach – which results only in a convex metatechnology

$T^M = \text{convex}\{T^1 \cup T^2 \cup \dots \cup T^G\}$ – suffers from a serious drawback: the estimated metafrontier may not envelop the estimated group frontiers so that $MTR(\mathbf{x}_j^g, y_j^g) = \text{eff}_j^M / \text{eff}_j^g > 1$ for some j and g .

In order to overcome this inconsistency, we first eliminate within-group inefficiencies from the data by replacing the observed output level y_j^g with $\hat{y}_j^g = y_j^g / \text{eff}_j^g$. We now suggest that the following variant of model (2) be solved for the estimation of the shape of function $F^M(\mathbf{x})$ such that $\hat{y}_j^g = F^M(\mathbf{x}_j^g) + \varepsilon_j^g$ and $\varepsilon_j^g \leq 0$:

$$\begin{aligned} \min \quad & \sum_{g=1}^G \sum_{j=1}^{\delta_g} (\varepsilon_j^g)^2 \\ \text{s.t.} \quad & \hat{y}_j^g = \alpha_j^g + \beta_j^g \mathbf{x}_j^g + \varepsilon_j^g, & \forall j, \forall g \\ & \alpha_j^g + \beta_j^g \mathbf{x}_j^g \leq \alpha_h^g + \beta_h^g \mathbf{x}_j^g, & \forall j, \forall g, \forall h, \forall q \\ & \beta_j^g \geq 0, & \forall j, \forall g \\ & \varepsilon_j^g \leq 0, & \forall j. \end{aligned} \quad (3)$$

This model computes inefficiency residuals ε_j^g , where $\hat{y}_j^g - \varepsilon_j^g = F^M(\mathbf{x}_j^g)$. This leads to $\hat{y}_j^g / F^M(\mathbf{x}_j^g) = (y_j^g / \text{eff}_j^g) / (y_j^g / \text{eff}_j^M) = \text{eff}_j^M / \text{eff}_j^g$, meaning that the model in (3) determines directly the metatechnology ratio of each firm, i.e. $MTR(\mathbf{x}_j^g, y_j^g) = \hat{y}_j^g / F^M(\mathbf{x}_j^g)$. Having computed the within-group efficiencies eff_j^g by the method in Section 3, the meta-efficiencies can then be calculated by $\text{eff}_j^M = MTR(\mathbf{x}_j^g, y_j^g) \times \text{eff}_j^g$. We note that as $\varepsilon_j^g \leq 0$, then $\hat{y}_j^g \leq F^M(\mathbf{x}_j^g)$. This implies that we will always have $MTR(\mathbf{x}_j^g, y_j^g) \leq 1$ and $\text{eff}_j^M \leq \text{eff}_j^g$.

Consider the case where the metatechnology is non-convex, i.e. $T^M = T^1 \cup T^2 \cup \dots \cup T^G$. The observed output level y_j^g is replaced by $\hat{y}_j^g = y_j^g / \text{eff}_j^g$. In order to estimate the shape of function F^M for this scenario, we suggest the following model be solved first for each group $g \in \mathbf{G}$:

$$\begin{aligned}
& \min \sum_{j=1}^{\delta_g} (\varepsilon_j^g)^2 \\
& \text{s.t. } \beta_j^g = \alpha_j^g + \beta_j'^g \mathbf{x}_j^g + \varepsilon_j^g, \quad \forall j \\
& \quad \alpha_j^g + \beta_j'^g \mathbf{x}_j^g \leq \alpha_h^g + \beta_h'^g \mathbf{x}_j^g, \quad \forall j, \forall h \\
& \quad \beta_j'^g \geq 0, \quad \forall j \\
& \quad \varepsilon_j^g \leq 0, \quad \forall j.
\end{aligned} \tag{4}$$

This model – as a modification of (2) – characterizes functions F^g in group g so that $\beta_j^g - \varepsilon_j^g = F^g(\mathbf{x}_j^g)$ and $\varepsilon_j^g \leq 0$. According to Kuosmanen and Johnson (2010), given the estimated coefficients α_j^g and $\beta_j'^g$ from (4), an explicit estimator of each function F^g can be constructed as $F^g(\mathbf{x}) = \min \{ \alpha_j^g + \beta_j'^g \mathbf{x}, j = 1, \dots, \delta_g \}$. With respect to the metatechnology $T^M = T^1 \cup T^2 \cup \dots \cup T^G$, determining $F^M(\mathbf{x})$ is equivalent now to finding the maximum of $F^g(\mathbf{x})$ across all groups as follows:

$$F^M(\mathbf{x}) = \max \left\{ \min \{ \alpha_j^g + \beta_j'^g \mathbf{x}, j = 1, \dots, \delta_g \}, g = 1, \dots, G \right\}. \tag{5}$$

Formula (5) gives an explicit estimator of function $F^M(\mathbf{x})$. Having computed the within-group efficiencies eff_j^g by the method in Section 3, we can then compute the meta-efficiencies and the meta-technology ratios by means of (5) and the relation $MTR(\mathbf{x}_j^g, y_j^g) = \beta_j^g / F^M(\mathbf{x}_j^g) = eff_j^M / eff_j^g$.

5. More on computation of meta-efficiencies

Consider a sign-constrained variant of the QP model in (2) in the sense that $\varepsilon_j^g \leq 0$. Kuosmanen and Johnson (2010) have proven that residuals $\varepsilon_j^g \leq 0$ in such a model can equivalently be computed by an appropriate DEA estimator (see Theorem 3.1 on page 152 of their paper). Within the convex metafrontier analysis, this means that each particular residual $\varepsilon_p^q \leq 0$ in model (3) for a firm p which belongs to group q could be calculated by the following linear problem:

$$\begin{aligned}
\hat{\varphi}_p^q &= \max \varphi \\
\text{s.t. } & \sum_{g=1}^G \sum_{j=1}^{\delta_g} \lambda_j^g y_p^g \geq \varphi y_p^q, \\
& \sum_{g=1}^G \sum_{j=1}^{\delta_g} \lambda_j^g x_{ij}^g \leq x_{ip}^q, \quad \forall i \\
& \sum_{g=1}^G \sum_{j=1}^{\delta_g} \lambda_j^g = 1, \\
& \lambda_j^g \geq 0, \quad \forall g, \forall j,
\end{aligned} \tag{6}$$

where $\varepsilon_p^q = y_p^q(1 - \hat{\varphi}_p^q)$ and $y_p^q - \varepsilon_p^q = F^M(\mathbf{x}_p^q)$. These two together imply that $(\hat{\varphi}_p^q)^{-1} = y_p^q / F^M(\mathbf{x}_p^q)$. As shown earlier in Section 4, we have $y_p^q / F^M(\mathbf{x}_p^q) = MTR(\mathbf{x}_p^q, y_p^q) = eff_p^M / eff_p^q$. Hence, the model in (6) computes the meta-technology ratios in the sense that $(\hat{\varphi}_p^q)^{-1} = eff_p^M / eff_p^q$.

For the non-convex metafrontier, the following shows that the same task of computing a meta-technology ratio by (4) and (5) together can be accomplished by solving the single mixed-integer linear program below:

$$\begin{aligned}
\hat{\varphi}_p^q &= \max \varphi \\
\text{s.t. } & \sum_{j=1}^{\delta_g} \lambda_j^g y_p^g \geq \varphi y_p^q - M\chi_g, \quad \forall g \\
& \sum_{j=1}^{\delta_g} \lambda_j^g x_{ij}^g \leq x_{ip}^q + M\chi_g, \quad \forall g, \forall i \\
& \sum_{j=1}^{\delta_g} \lambda_j^g = 1, \quad \forall g \\
& \sum_{g=1}^G \chi_g = G - 1, \\
& \chi_g \in \{0, 1\}, \quad \lambda_j^g \geq 0, \quad \forall g, \forall j.
\end{aligned} \tag{7}$$

All equivalent linear estimators of the models in (4) have been incorporated into model (7) by the first three constraints. However, binary variables χ_g , $g \in \mathbf{G}$, as well as a sufficiently large constant M ensure that only one of the groups will be selected at any time. This group will be the one by which the

objective function is maximized. This implies that the binary procedure here accomplishes the same task as the formula in (5). Hence, the optimal objective function value of (7) computes the meta-technology ratios such that $(\hat{\phi}_p^q)^{-1} = MTR(\mathbf{x}_p^q, y_p^q) = eff_p^M / eff_p^q$.

Acknowledgement

The financial support of the Deutsche Forschungsgemeinschaft (DFG) in the context of the research fund AH 90/5-1 is gratefully acknowledged.

References

- Aigner, D., Lovell, C.A.K., Schmidt, P. 1977. Formulation and estimation of stochastic frontier production function models. *J. Econom.* 6(1), 21-37.
- Battese, G.E., Rao, D.P., O'Donnell, C.J. 2004. A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. *J. Prod. Anal.* 21(1), 91-103.
- Hayami, Y., Ruttan, V.W. 1970. Agricultural productivity differences among countries. *American Econ. Rev.* 60(5), 895-911.
- Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P. 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *J. Econom.* 19(2-3), 233-238.
- Kerstens, K., O'Donnell, C.J., Van de Woestyne, I. Frontier metatechnologies and convexity: A restatement, working paper 2015-08. IESEG School of Management, France, 2015.
- Kuosmanen, T. 2006. Stochastic nonparametric envelopment of data: Combining virtues of SFA and DEA in a unified framework. MTT Discussion Paper No. 3/2006.
- Kuosmanen, T., Johnson, A.L. 2010. Data envelopment analysis as nonparametric least-squares regression. *Oper. Res.* 58(1), 149-160.
- Kuosmanen, T., Johnson, A., Saastamoinen, A. 2015. Stochastic nonparametric approach to efficiency analysis: A unified framework. In *Data Envelopment Analysis* (pp. 191-244). Springer US.
- O'Donnell, C.J., Rao, D.P., Battese, G.E. 2008. Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Emp. Econom.* 34(2), 231-255.
- Sipiläinen, T., Kuosmanen, T., Kumbhakar, S.C. 2008. Measuring productivity differentials – An application to milk production in Nordic countries. 12th Congress of the European Association of Agricultural Economics, Ghent, Belgium.